HEP-CCE

Event/Data Batching for DUNE

Kyle Knoepfel (FNAL), Meifeng Lin (BNL), Haiwang Yu (BNL), Brett Viren (BNL)

HEP-CCE PPS Meeting
March 19, 2021

U.S. DEPARTMENT OF ENERGY | Office of Science
Argonne NATIONAL LABORATORY
BROOKHAVEN NATIONAL LABORATORY
Fermilab
BERKELEY LAB Bringing Science Solutions to the World

Event/data batching why and what

- ***Why:*** To increase CPU/GPU utilization and reduce data transfer overhead
- **Three potential granularity levels of "batching" in DUNE/WCT**
  - ***Fine:*** In the same algorithm, bundle many energy depos into a single batch.
    - Already doing this in wire-cell-gen-kokkos as current PPS activity
  - ***Coarse:*** Do we gain by batching simulations of different APAs?
    - Depends on the activity level.
    - May need to select "interesting" APAs to batch.
  - ***Global:*** How do we decide when GPU is needed and when CPU is preferred?
    - Only use GPU when gain is more than CPU alone
    - Need fair comparison between multithreaded CPU vs. GPU
    - Can we batch different GPU-enabled algorithms together?

Effort Estimate

- Except for the “fine”-level data batching, it is unclear how much gain we get from event/data batching in DUNE/WCT in general.
- Will need some effort to evaluate what is done currently and what can be done differently to improve the CPU+GPU throughput.
  - Goal is to find use cases that can benefit from event batching such that CPU+GPU throughput is greater than multithreaded CPU throughput.
- Likely use case: APA batching
- Effort Estimate:
  - Evaluation of potential gain from event batching: ~ **1 FTE*month**
  - Implementation of the example event batching: ~ **2 FTE*month**
    - **Contingent on the evaluation**
  - Performance evaluation and recommendation: ~ **1 FTE*month**
    - **Contingent on the evaluation**